Sparse-view phase-contrast and attenuation-based CT reconstruction utilizing model-driven deep learning*

Xia-Yu Tao,[1] Qi-Si Lin,[1] Zhao Wu,[1, †] Yong Guan,[1] Yang-Chao Tian,[1] and Gang Liu[1, ‡]

[1]*National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, 230009, China*

Grating-based X-ray phase-contrast imaging enhances the contrast of imaged objects, particularly soft tissues. However, the radiation dose in computed tomography (CT) is generally excessive owing to the complex collection scheme. Sparse-view CT collection reduces the radiation dose, but with reduced resolution and reconstructed artifacts particularly in analytical reconstruction methods. Recently, deep learning has been employed in sparse-view CT reconstruction and achieved state-of-the-art results. Nevertheless, its low generalization performance and requirement for abundant training datasets have hindered the practical application of deep learning in phase-contrast CT. In this study, a CT model was used to generate a substantial number of simulated training datasets, thereby circumventing the need for experimental datasets. By training a network with simulated training datasets, the proposed method achieves high generalization performance in attenuation-based CT and phase-contrast CT, despite the lack of sufficient experimental datasets. In experiments utilizing only half of the CT data, our proposed method obtained an image quality comparable to that of the filtered back-projection algorithm with full-view projection. The proposed method simultaneously addresses two challenges in phase-contrast three-dimensional imaging, namely, the lack of experimental datasets and the high exposure dose, through model-driven deep learning. This method significantly accelerates the practical application of phase-contrast CT.

Keywords: Sparse-view CT, Phase-contrast CT, Attenuation-based CT, Deep learning network, Frequency loss function

I. INTRODUCTION

Computed tomography (CT) has become an indispensable imaging tool in clinical practice [1]. CT contributes to the noninvasive and painless diagnosis of human organs of interest [2] and crucial in preoperative evaluation and treatment planning [3]. Medical CT is generally based on the absorption principle; however, the low contrast of soft tissues hinders the early diagnosis of cancer and other diseases [4]. Grating-based X-ray phase-contrast CT offers multicontrast and enhanced contrast for low-Z soft tissues and provides the possibility of early diagnosis [5–8]. Regrettably, CT requires an extra X-ray dose, which can be damaging to patients [9], particularly their DNA [10–12]. In phase-contrast CT, the radiation dose is several times higher than that in conventional CT because it requires multiple projections at each tomographic viewing angle to retrieve multicontrast information [13]. Reducing the dose generally leads to lower image quality and potential misdiagnosis [12]. However, low-dose imaging methods have been proposed to maintain the signal-to-noise ratio (SNR) by leveraging prior knowledge. Therefore, achieving a balance between effective medical examinations and minimizing radiation damage is crucial [11].

There are two approaches to decreasing X-ray radiation damage: low-dose and sparse-view CT. In low-dose CT, the X-ray exposure in each view is reduced, and a photon-counting detector can be utilized to maintain the SNR of the projections. In sparse-view CT, violations of the Shannon–Nyquist sampling theorem lead to reduced resolution, artifacts, and distortions in the reconstructed image [14]. This study focuses on sparse-view CT and introduces a new reconstruction algorithm aimed at suppressing these artifacts and distortions, thereby enhancing image quality.

Filtered back projection (FBP) is a method widely used in modern CT systems for high-dose full-view CT because it provides rapid and high-quality results with minimal computational resources [15, 16]. However, when applied to sparse-view CT, FBP often generates significant stripe artifacts. Iterative reconstruction (IR) algorithms, such as the simultaneous algebraic reconstruction technique (SART) [17] and simultaneous iterative reconstruction technique (SIRT) [18], can partially suppress artifacts through iterative forward projection and backward correction based on convex optimization theory [19]. With the advancement of the compressed sensing (CS) algorithm, which enables signal reconstruction from undersampled data [20], numerous studies have focused on the total variation (TV) model, which utilizes the gradient of parametric L1 to smooth images [23, 24]. TV methods serve as regularization terms in the cost functions of IR algorithms by incorporating prior knowledge [25]. In certain ideal scenarios, TV models, such as TV-Projection Onto Convex Sets (POCS) [26], SART-TV [23], and Total Generalized Variation Regularization (TGV) [25] can effectively eliminate stripe artifacts in the reconstructed images [27].

Recently, deep learning (DL) has been widely adopted for various image processing tasks, including image denoising [31, 32], image recognition [33], image segmentation [34], image inpainting [35], and image super-resolution [36]. DL-based algorithms have shown remarkable performance improvements over traditional methods, particularly in handling noisy images and enhancing the image quality [37, 38]. Researchers have also applied DL technol-

ogy in sparse-view CT and investigated the significance of datasets. Hwan et al. proposed an end-to-end deep convolutional network-based U-net (FBPConvNet) that was trained using reconstructed slices from both sparse-view and full-view CT scans as the input and output, respectively, [39]. They utilized a biomedical dataset for training and achieved superior results compared with the TV method. Han et al. demonstrated that DL networks could effectively distinguish streaking artifacts from artifact-free images [40]. They employed a deep residual-learning architecture trained on data from nine patients to suppress streaking artifacts. Han et al. highlighted the limitations of the U-Net architecture, which excessively emphasized the low-frequency components of the signal, resulting in blurred image edges [41]. To address this issue, they proposed a new multiresolution DL framework to recover high-frequency edges in sparse-view CT. Guan et al. introduced the fully dense U-Net (FD-UNet) to remove artifacts in 2D-PAT (photoacoustic tomography) images reconstructed from sparse data [42]. However, they observed that the performance of FD-UNet deteriorated when the training and testing data did not effectively match. Asif et al. utilized a GAN to generate cardiac images and suppress cardiac motion artifacts [43]. They also proposed diffusion and score-matching models for generating CT images from MRI images [44]. Nevertheless, the current DL-based sparse-view CT reconstruction algorithms rely heavily on experimental datasets. Acquiring sample datasets such as medical datasets is a laborious, time-consuming, and expensive process. Moreover, the limited data do not guarantee the reliability of DL algorithms. In phase-contrast imaging experiments, the test samples are varied, and obtaining numerous full-view datasets in advance is not consistently feasible. Consequently, alternative datasets are required for training.

Previous studies have shown that natural and medical images share common low-level features and similarities in terms of edges, points, and textures [45]. The transfer of prior knowledge from natural image processing to medical image processing has been validated in several studies. For instance, Zhong et al. synthesized the noise in natural images for low-dose CT (LDCT) denoising and transferred the learned knowledge to medical images to prevent overfitting during training [46]. In another study, Zhen et al. pretrained a classification network on ImageNet and fine-tuned a convolutional neural network (CNN) for transfer learning to predict the toxicity of a cervical cancer rectal dose [47]. These studies demonstrated similarities between natural and medical images in terms of pixel correlation and low-level features. Consequently, natural image datasets are excellent for deep learning reconstruction of phase-contrast images without training data.

Motivated by these studies, we propose a physical model of limited-angle CT that utilizes natural images to generate abundant high-quality data [48]. Excellent reconstruction results were achieved by incorporating an optimized network structure and loss function.

In this study, model-driven DL was introduced to solve the issues of limited experimental training datasets and high exposure doses in sparse-view phase-contrast CT. The CT device was parameterized for both attenuation-based and phase-contrast CT procedures, allowing the generation of simulation datasets. The reconstruction results of sparse-view attenuation-based CT and phase-contrast CT demonstrate that the proposed method substantially suppresses artifacts.

The main contributions of this study are summarized as follows.

1. We propose a novel DL CT reconstruction method that integrates an X-ray phase-contrast imaging model with superior generalization capabilities. By eliminating the network's dependence on the experimental data, our method improves the accuracy and robustness of sparse-view CT reconstruction.

2. Furthermore, a new frequency loss function is introduced based on the Fourier Slice Theorem. This loss function transforms the projection data into a fidelity term of the network via a Fourier transform, resulting in enhanced image generation quality.

3. Superior performance compared to traditional algorithms was realized using experimental data from both attenuation-based CT and phase-contrast CT. This improved performance highlights the potential of our method for advanced applications of phase-contrast CT.

The remainder of this paper is organized as follows: Section II introduces and discusses the proposed algorithm and its detailed framework. Section III presents experimental results obtained by applying the proposed method to a laboratory phase-contrast CT. Section IV discusses the strengths and limitations of the proposed algorithm. Finally, section V summarizes the findings.

II. METHODS

Figure 1 illustrates the architecture of the proposed sparse-view reconstruction framework based on model-driven simulated big data. We acquired natural images from the Common Objects in Context (COCO) 2017 dataset [49], which consists of various animals, scenery, architecture, food, and more. The first step of the proposed method involves image batch preprocessing, which standardizes the sizes of all images and eliminates grayscale variations. Subsequently, sparse-view CT data were simulated from the preprocessed normalized images using a forward projection algorithm specifically designed for grating-based X-ray phase-contrast CT equipment. Three-dimensional images with artifacts were reconstructed using the FBP algorithm. These reconstructed images were fed into an end-to-end DL network based on Unet. The normalized images served as the ground truth (GT) to train the network parameters and were evaluated using a loss function. Finally, a DL network was employed to obtain artifact-free reconstructed images by applying FBP to images with artifacts.
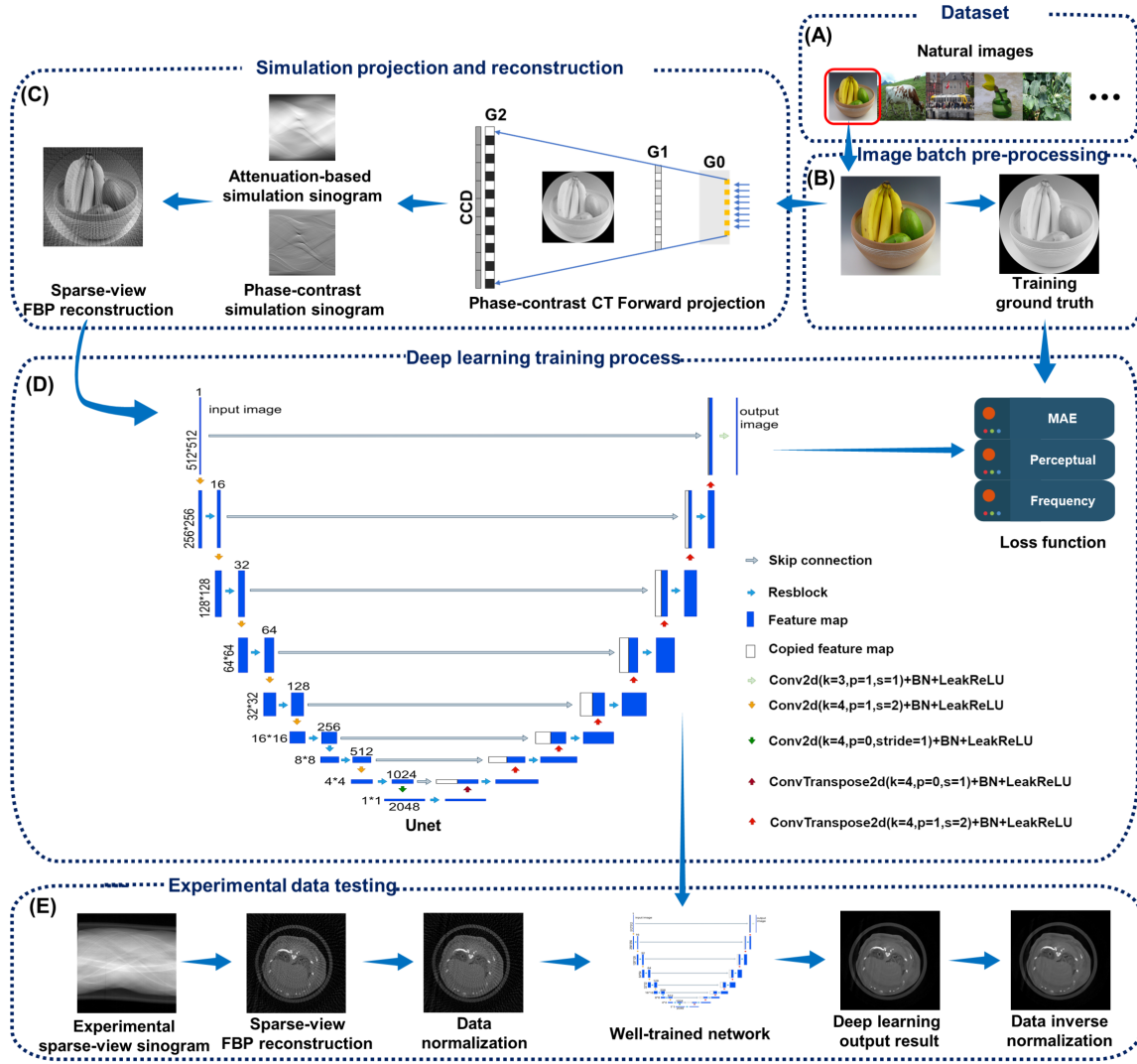
Fig. 1. (Color online) The framework of the proposed sparse-view reconstruction based on model-driven simulated big data. The process is as follows: (a) The dataset consists of raw data from natural images. (b) The images undergo pre-processing and normalization. (c) The normalized images are forward-projected and reconstructed using the FBP algorithm. (d) The network architecture and its training are based on simulated data. (e) The experimental artifact-free images are obtained by feeding the FBP-reconstructed images into a DL network trained on the simulated dataset.

A. Data set

In this study, a CT model was used to generate a simulated natural dataset rather than experimental CT projection data. Natural data were obtained from the COCO2017 dataset, which contained a diverse collection of 118,287 images featuring animals, scenery, architecture, and food. The natural dataset exhibited a rich image distribution, facilitating generalization to unknown sample data. In the results section, we utilize the training data from the natural dataset for our proposed method. However, dissimilarities arose in the final training data employed for attenuation-based CT and phase-contrast CT because of disparities in their respective CT projection models. Consequently, the sparse-view recon-

struction images obtained through CT model simulations exhibited variations between the two modalities.

A medical dataset was used for comparison. The medical dataset was sourced from the American Association of Physicists in Medicine (AAPM) Low Dose CT Grand Challenge and the Cancer Imaging Archive (TCIA) "Low Dose CT Image and Projection Data" dataset [58]. The medical dataset was simulated and derived from full-dose lung CT images from the AAPM dataset. A total of 5,623 images were obtained from the data of 30 patients and through data augmentation techniques such as image rotation, a final dataset of 118,287 images was generated. The calculation methods employed for the medical dataset were consistent with those used for the proposed natural dataset, except for image incon-

B. Image batch pre-processing

Because of the input size limitation of the DL network, the images were resized to a uniform size of 512×512 pixels. The resizing process involves interpolation, scaling, and cropping. Subsequently, the images were converted to grayscale and normalized. During training, the GT image for the network was generated using circular artifact-free images obtained via image-batch preprocessing.

C. Attenuation-based projection and CT reconstruction

In the calculation of attenuation-based CT projections, the relationship between the object image function and the projection data in each view can be expressed using the following equation:

$$P_\theta = A_\theta X \tag{1}$$

In the discrete model, $P_\theta = (p_i) \in R^M$ represents a one-dimensional column vector array that captures the line integral of an object image at projection angle θ, where M is the number of detectors. As each projection value may be influenced by all values in the object image, we arrange the two-dimensional image into a one-dimensional column vector: $X = (x_j) \in R^N$ where x_j represents each pixel in X, and N represents the total number of pixels. $A_\theta = (a_{ij}) \in R^{M*N}$ represents a two-dimensional projection system matrix of size $M * N$, where a_{ij} represents the coefficient of contribution of pixel point x_j in the object image to the line integral projection value p_i.

To reduce the errors in the discrete calculation, we adopted an area projection model, where a_{ij} represents the ratio of the overlapping area of ray i with voxel j to the area of the pixel [51]. This study primarily focused on the geometry of the parallel beam and employed the 'fan2para' function in MATLAB to convert the experimental data from fan-beam geometry to the parallel case.

Sparse-view CT reconstructions were generated using 90 and 45 projection views via the FBP algorithm with a ramp filter kernel. During the network training process, we used sparse-view CT reconstruction results as inputs and object images as outputs to train the network and adjust the network parameters using the loss function.

D. Phase-contrast projection and CT reconstruction

Compared with attenuation-based projection, phase-contrast projection requires additional differential calculations in the direction perpendicular to the optical and rotation axes. In CT reconstruction, the filter kernel is a Hilbert function that directly retrieves the decrement of the refractive index without requiring integration [53].

E. Deep learning network structure

The proposed DL network was optimized from the Unet architecture, incorporating various layers such as convolutional layers (Conv2d), batch normalization layers (BN), layers with the Leaky ReLU activation function, deconvolutional layers (ConvTranspose2d), residual blocks (ResBlock) [54], and a tanh layer to output the result. To enhance the network performance, additional downsampling layers were introduced in the network structure to expand the receptive field and capture more high-level information. To prevent information loss, the pooling layers were replaced with convolutional layers that utilized larger strides. Moreover, residual blocks with expandable depths were utilized to further increase network capacity and improve feature extraction. In the skip connection, the features are directly concatenated with the corresponding features in the upsampling path to retain crucial low-level information. Using the following calculation, the network can be designed based on the dimensions of the input and output images.

Convolutional layers can extract information from the input images, and weights can be shared across the network. Using convolutional kernels of different sizes and strides, various feature sampling functions can be achieved. The sizes of convolutional input and output were calculated as follows:

$$N = \frac{(W - F + 2P)}{S} + 1 \tag{2}$$

where W represents the size of the input feature map, N is the size of the output feature map, F is the size of the convolution kernel, P denotes the padding size that fills the feature map with zero values, and S denotes the convolutional stride. The network does not include any pooling layers, as they may lead to the loss of significant information during training. Rather, we employed a convolutional kernel with a stride size of 2 to achieve the desired effect. The effective convolution size of the network can be increased by incorporating feature maps with different resolutions.

Batch normalization layers are effective in accelerating convergence and preventing gradient explosions. They also serve as regularization techniques to mitigate network overfitting. Leaky ReLU layers are nonlinear activation functions that enable a network to perform effectively in nonlinear computations. The mathematical expression for the Leaky ReLU is as follows:

$$y = \max(0, x) + \text{leak} * \min(0, x) \tag{3}$$

Here, the leak is a small constant that retains information from the negative axis.

Deconvolutional layers are utilized to restore the feature map size and facilitate feature extraction and are primarily employed in image generation. The sizes of deconvolutional input and output are calculated as follows:

$$N = (W - 1) * S - 2P + F \tag{4}$$

Shortcut connections also referred to as skip connections, allow the direct transfer of features from different layers to subsequent layers, thus preventing the loss of information from deeper features.

ResBlock is a convolutional layer that maintains a consistent feature-map size. It increases the network parameters without overly increasing its susceptibility to overfitting, thereby improving the overall performance.

The tanh layer serves as the output layer, generating images with pixel values ranging from 0 to 1. The mathematical expression for the tanh layer is as follows.

$$y = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

F. Loss function

A mixed loss function was designed by combining mean absolute error (MAE), perceptual loss [55], and frequency loss. The loss function is expressed as follows:

$$\begin{aligned} Loss = {} & \alpha * \|Y - f(X)\| \\ & + \beta * PreceptualLoss \\ & + \gamma * FrequencyLoss \end{aligned} \tag{6}$$

where Y represents the GT images, X represents the images input to the network, $f(X)$ represents the output images from the network, and $\|Y - f(X)\|$ represents the MAE. The scale factors α, β and γ are used to adjust the contribution of each loss component.

MAE loss primarily aims to prevent image distortion, making the network training output resemble that of the GT images. Compared with the mean squared error (MSE) loss ($\|(Y - f(X))^2\|$), the MAE produces images generated by the DL network with sharper edges. However, MAE may produce special singularities with large error values, necessitating the incorporation of additional loss functions to mitigate pixel-generation errors.

Perceptual Loss enhances the accuracy of the generated images by considering low- and high-level features. For this purpose, a pretrained VGG19 network was utilized as a part of the loss function. The VGG19 network was trained on the ImageNet dataset and its parameters were fixed after training. VGG19 is a CNN capable of extracting high- and low-level features at different resolutions. Low-level features are typically found in the early convolutional layers, whereas high-level features are typically present in the later convolutional layers. The mathematical expression for perceptual loss is as follows:

$$Perceptual\ Loss_{\text{feature}}^{\phi, j} = \frac{1}{C_i H_i W_i} \|\phi_j(\widehat{y}) - \phi_j(y)\|_2^2 \tag{7}$$

Here, j represents the number of network layers of VGG16; $C_j H_j W_j$ represents the size of the feature map in layer j; y and $\widehat{y}$ represent the GT and network output, respectively; and $\phi_j(x)$ represents the output of the feature map in layer j.

The presence of stripe artifacts can be attributed to the loss of information in the frequency spectrum domain, according to the Fourier Slice Theorem. The Fourier-transformed image distinguishes between high- and low-frequency information within the image. According to Parseval's theorem, the MAE loss in the spatial domain is equivalent to the MAE loss in the frequency spectrum domain. In the spectral domain, the low-frequency information coefficients are an order of magnitude higher than the high-frequency information, which explains the smoother appearance of the MAE loss images. The frequency loss component also helps reduce the occurrence of checkerboard artifacts.

To mitigate the impact of low-frequency components and focus network learning on high-frequency information, an adaptive frequency spectrum domain loss based on the MAE loss was employed. Assume that the DL network generates images as $f(x, y)$ with a two-dimensional Fourier transform of $F(u, v) = a + bi$ and that the GT generates images as $f_0(x, y)$ with a two-dimensional Fourier transform of $F_0(u, v) = a_0 + b_0 i$. The mathematical expression for the frequency-spectrum domain loss is as follows:

$$Frequency\ Loss = mean(w * (a - a_0) + w * (b - b_0)) \tag{8}$$

$$w = log((a - a_0) + (b - b_0) + 1) \tag{9}$$

where w denotes the dynamic update factor. During network training, w reduces the overlearning of low-frequency parts and emphasizes the learning of high-frequency information.

G. Experimental data testing

For experimental data testing, we normalized the FBP reconstruction results because of the limited range of values for the network input and output. In addition, after obtaining the de-artifacting output from the network, we must apply inverse normalization to the image data. The final computed image represents the reconstruction results obtained using the proposed method.

H. Image evaluation criteria

We evaluated the performance of the proposed algorithm using two metrics: Peak SNR(PSNR) and structural similarity index (SSIM). PSNR measures the ratio between the maximum possible power of an image and the power of the corrupting noise that affects the quality of its representation. The PSNR formula is as follows:

$$PSNR = 10\log_{10}\frac{(L-1)^2}{MSE} \qquad (10)$$

where L denotes the number of maximum possible intensity levels (with the minimum intensity level assumed to be zero) in the image.

MSE is defined as follows:

$$MSE = \frac{1}{MN} \sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(O(i,j) - D(i,j))^2 \qquad (11)$$

In the above equation, O represents the original image, D represents the reconstructed image, m represents the number of rows of pixels, i represents the index of that row of the image, n represents the number of columns of pixels, and j represents the index of that column of the image.

SSIM, however, measures the structural similarity between the original and reconstructed images. The SSIM formula is defined as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x{}^2 + \mu_y{}^2 + C_1)(\sigma_x{}^2 + \sigma_y{}^2 + c_2)} \qquad (12)$$

where y represents the original image, x is the reconstructed image, μ_x is the average value of x, μ_y is the average value of y, $\mu_x{}^2$ is the variance of x, $\mu_y{}^2$ is the variance of y, σ_{xy} is the covariance of x and y, c_1 and c_2 are the constants. Specifically, $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$, where L is the dynamic range of the pixel values, k_1=0.01,and k_2=0.03.

III. RESULTS

First, in this section, we outline the key steps involved in training the proposed DL network and provide details regarding the computing environment. Subsequently, the reconstruction efficiency of the proposed algorithm was demonstrated using attenuation CT and phase-contrast CT. Finally, a comparison was made between the generalization performance of training using natural data and that of training using medical data.

A. Network training details

The proposed DL network is trained using the Adam algorithm [56]. The learning rate was set to 0.0001 and the mini-batch size was 8. In the loss function, α and β were both set to 1, whereas γ was set to 0 for the first 10 epochs. This configuration allowed the network to learn the main features and converge faster. After 10 epochs, α was set to 0, β was set to 0, and γ was set to 1, which ensured that the network focused more on high-frequency information. The proposed network is implemented using PyTorch [57] on a personal workstation equipped with an Intel(R) Core (TM) i9-10940X CPU and

128GB of RAM. NVIDIA GeForce RTX 3090Ti was used to accelerate the network training operations.

For the X-ray phase-contrast CT of mice, a grating-based phase-contrast imaging system was employed for X-ray phase-contrast CT of mice. A step-by-step phase method was employed to acquire attenuation-based and phase-contrast sinograms. Experimental results for both modes are presented below to evaluate the performance of the proposed method.

B. Experimental results of attenuation-based CT in phase-contrast imaging

Figure 2 illustrates the reconstructed lung slices of mice using various methods, including FBP, SART, SART-TV algorithms, and our proposed method. GT was defined as a 180-views SART-TV reconstruction. Additionally, a phase-contrast slice with enhanced contrast was included for structural comparison. These methods were employed to reconstruct slices using both 2-degree interval projections (90 views) and 4-degree interval projection (45 views) interval projections. The residual images are the error images between the reconstruction results of the different algorithms and the GT.

The FBP and SART algorithms performed poorly in both cases. In a scenario with 90 views, the SART-TV algorithm effectively suppressed stripe artifacts caused by sparse-view projections. However, noticeable smoothness and blocky artifacts were observed in the zoomed-in inner edges. When there were only 45 views, the SART-TV algorithm performed poorly, displaying excessively smooth local details and prominent streaking artifacts in the overall image. In contrast, the proposed algorithm successfully removed artifacts and maintained consistency with the GT slices in both cases. It also avoids generating blocky artifacts observed in the SART-TV algorithm. For the residual images, the proposed algorithm exhibits the smallest error. When using the phase-contrast results as references, the local zoomed-in structure is even clearer in the 90-view case using the proposed method than in the 180-view case using the SART-TV algorithm. Quantitative assessments of image quality showed that our proposed algorithm outperformed existing methods.

C. Experimental results of phase-contrast CT in phase-contrast imaging

Phase-contrast reconstruction of the laboratory phase-contrast CT equipment is shown in Figure 3. Currently, phase-contrast imaging algorithms rely primarily on the FBP algorithm, and we did not include a comparison with iterative algorithms. The phase-contrast projection image was calculated from the attenuation-based projection using the information separation method described in Section 2 (Methods). The GT results were reconstructed using the FBP algorithm with 180 projection views. In both the 90-view and 45 views scenarios, our proposed method outperformed the FBP algorithm in effectively removing streak artifacts caused
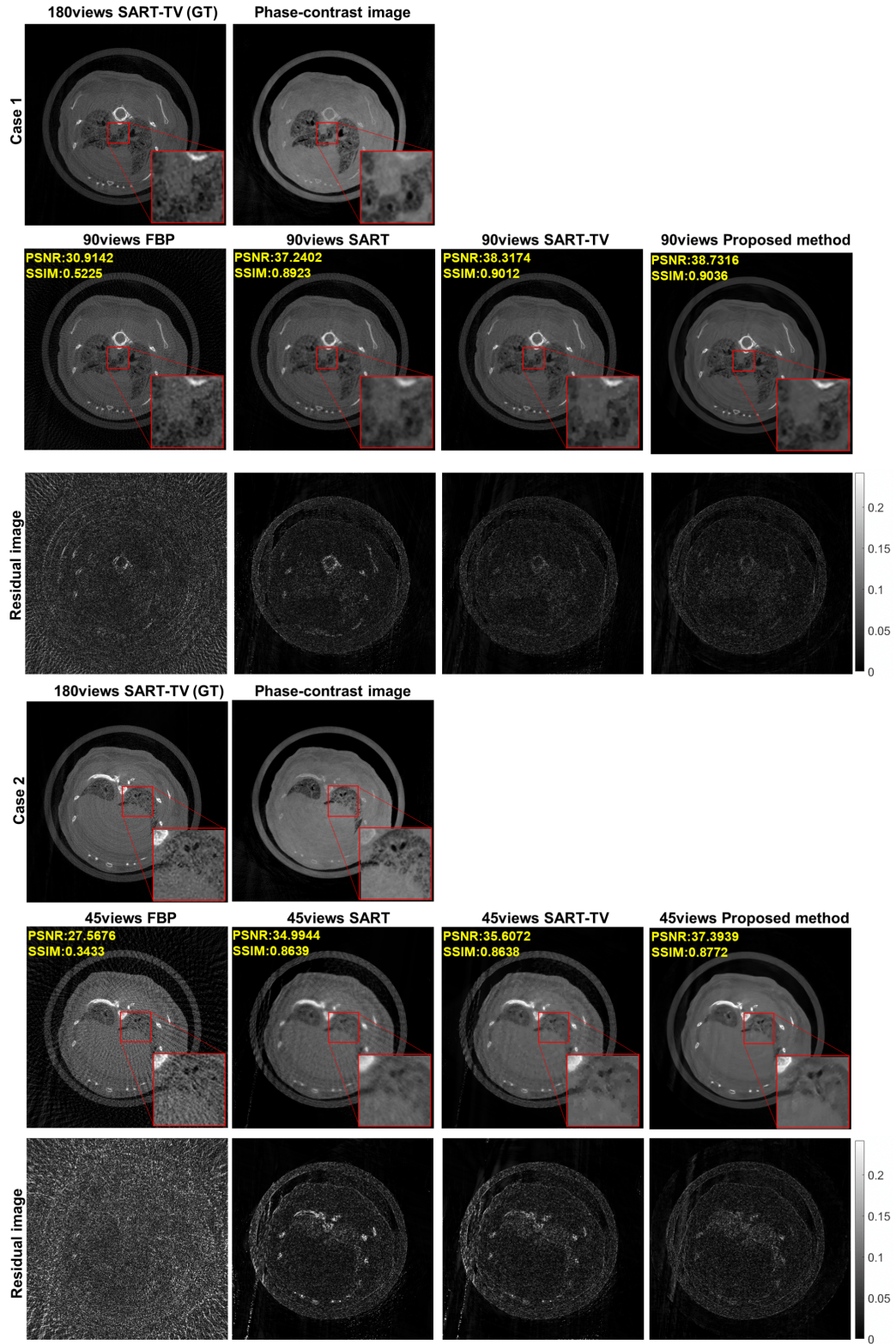
Fig. 2. (Color online) Comparison of reconstructed attenuation-based slices by different methods in phase-contrast imaging of a mouse. The proposed algorithm is compared with the conventional algorithms of FBP, SART, and SART-TV in sparse-view CT of 90 views and 45 views. The GT images were reconstructed using the SART-TV algorithm from 180 views projection. The SSIM and PSNR metrics with respect to the GT are provided in the top left corner of each panel. The regions of interest are magnified and indicated by the red boxes in the figure. The results also include residual images between different algorithms and the GT, where white pixels represent larger errors. The display range for all results is normalized to [0,255] HU.

by sparse-view projections and maintained consistency with the 180-view GT image. From the comparison of the residual images in Figure 3, the proposed method exhibits less error in the reconstruction results of the 90 views and 45 views projection data compared to the FBP algorithm. The evaluation metrics of the image quality also demonstrated the superior performance of our algorithm in reconstructing the results from sparse-view phase-contrast projections.

To validate the applicability of our method, we conducted additional calculations using the synchrotron radiation phase-contrast experimental data. Figure 3 shows the phase-contrast reconstruction obtained from the BL13W1 beamline at the Shanghai Synchrotron Radiation Facility (SSRF), China [59]. As this is a synchrotron light source, the projections are already in a parallel-beam geometry and do not require additional interpolation. The image shows a phase-contrast slice of a bee immersed in a microcentrifuge tube filled with formalin. The GT results were reconstructed using the FBP algorithm with 360-degree projection. Given the relatively simple structure of the bee, even with only 45 views, the proposed method achieved results comparable to those obtained with 360 views. The residual image comparison shows that the proposed method effectively removed the stripe artifacts.

The two sets of data presented in Figure 3 demonstrate the excellent performance of the proposed method in the field of grating-based phase-contrast imaging.

D. Experimental results under natural dataset and medical dataset

To verify the high generalization performance of the natural datasets, we compared the results generated by the networks trained on different datasets using attenuation-based CT and phase-contrast CT. This comparison is shown in Figure 4. For the attenuation-based CT, the GT image was defined as the reconstructed result obtained using the SART-TV algorithm from a 180-view projection. By contrast, in phase-contrast CT, the GT image is reconstructed using the FBP algorithm with 180-view projections. This selection was made because of the current prominence of the FBP algorithm in phase-contrast CT reconstruction, whereas the SART-TV algorithm is considered optimal for attenuation-based CT.

The attenuation-based CT and phase-contrast CT results are presented in Figure. 4 were obtained from the same mouse slice. Additionally, reconstructions were performed using 90 and 45 views to investigate the impact of data sparsity. The results shown in Figure 4 demonstrate that the network trained on the natural dataset outperformed that trained on the medical dataset for both attenuation-based CT and phase-contrast CT. The results obtained from the medical dataset exhibited a notably poorer performance, resulting in unclear and blurred local details. This discrepancy can be attributed to the limited image distribution of the medical dataset compared with the more diverse image distribution of the natural dataset. The greater diversity in the natural dataset contributes to its superior generalization performance when confronted with unknown samples. Furthermore, we con-

ducted simulation verification using enough medical clinical CT data. The results are provided in Supplementary Information (SI). These results demonstrate that natural data can achieve outcomes comparable to those of a sufficient amount of medical data, even without prior medical knowledge.

In attenuation-based CT, soft tissue contrast is low, and sparse sampling leads to a significant decrease in image resolution, resulting in blurred structures. Conversely, in phase-contrast CT, the soft-tissue contrast is high, and the structures remain remarkably clear, even with only 45 views. These findings further validate the promising prospects of the proposed algorithm for phase-contrast CT.

IV. DISCUSSION

Although the generalization performance of DL is currently not well understood, the experiments provided compelling evidence of the effectiveness of the proposed method in terms of generalization. Natural and phase-contrast sample images share common characteristics such as low-rank properties and similar low-level features such as points, lines, and edges. In traditional methods, the TV model is based on connections between pixels that are common to natural images. This explains the excellent generalization performance of natural image datasets in sparse-view CT reconstruction. These relationships warrant further investigation.

By modeling the imaging procedure, we can apply DL techniques with simulated training datasets. This approach is particularly valuable in scientific research and practical applications. Furthermore, the results obtained from our experiments demonstrate the effectiveness of the proposed method in both attenuation-based CT and phase-contrast CT. Notably, the proposed method maintains a high contrast in phase-contrast imaging and preserves clear structures even with a reduced dosage in sparse-view CT. Moreover, this physical model-based approach can be extended to other fields by adapting the parametric representation of the experimental equipment. In addition, the proposed method can be applied to other tasks involving artifact removal.

The current approach only considers samples within the field of view; reconstructing samples outside the field of view may require a reconsideration of the model design. Experimental results show that the proposed method can produce results comparable to those of the GT image in 90 views CT, but further optimization is required for complex samples in 45 views CT. Therefore, additional improvements are necessary to enhance reconstruction quality under severely sparse conditions, which may involve incorporating other inference models into the algorithm.

V. CONCLUSION

This study introduces a novel and promising approach that integrates a model-driven DL reconstruction algorithm into sparse-view phase-contrast three-dimensional imaging. This overcomes the limited availability of experimental training
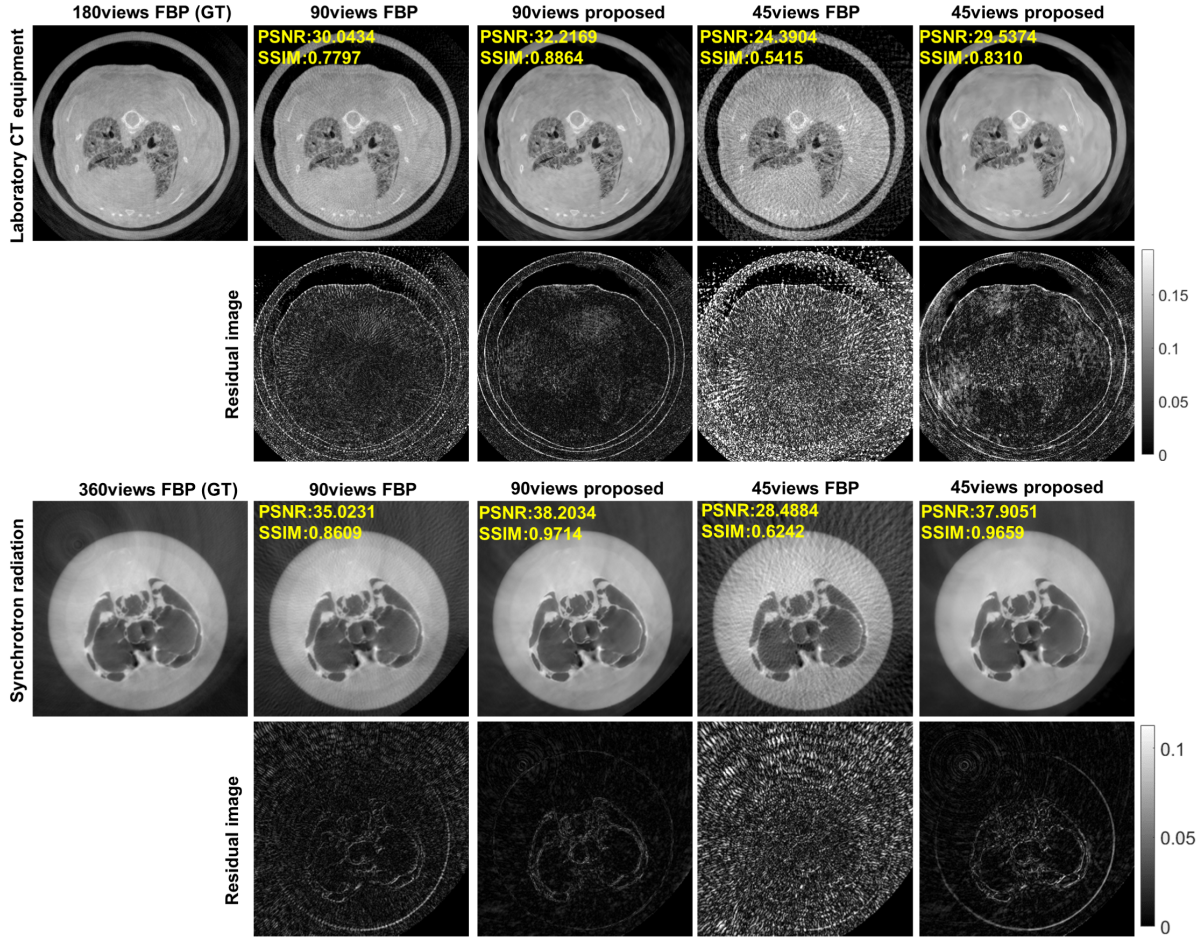
Fig. 3. (Color online) Comparison of the reconstructed slices of biological samples by the FBP algorithm and the proposed method in phase-contrast CT. The proposed algorithm is compared with the conventional algorithms of FBP in sparse-view CT scenarios with 90 views and 45 views. The GT image is reconstructed using the FBP algorithm with 180 and 360 views projection. The SSIM and PSNR metrics with respect to the GT are displayed in the top left corner of each panel. The results also include residual images between different algorithms and the GT, where white pixels represent larger errors. The display range for all results is normalized to [0,255] HU.

datasets. The experimental results demonstrate the superiority of the proposed method over conventional algorithms in terms of reconstruction quality. This effectively enhances the accuracy and fidelity of reconstructions in both sparse-view attenuation-based and phase-contrast CT. The reduced imaging time of the proposed method may enable in vivo phase-contrast imaging of biological specimens. This advancement opens new possibilities for applications in biological medicine, where the ability to capture high-resolution real-time images of living tissues and organs can provide valuable insights for diagnosis, treatment planning, and research. Furthermore, the proposed method offers a potential avenue for future research and development as well as potential clinical translation in the pursuit of more accurate and efficient imaging techniques.

Disclosures: The authors declare no conflicts of interest.

Data availability The data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Ethical approval The study protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of Precedo Pharmaceuticals Co., Ltd. During the study, the animal breeding and use will strictly follow the regulations of the International Management Committee for Assessment and Certification of Laboratory Animals (No. IACUC-20221201).

Contributions All authors contributed to the study's conception and design. Material preparation, data collection, and analyses were assisted by Xiayu Tao Qisi Lin and Zhao Wu. The coding of deep learning methods in the manuscript was assisted by Xiayu Tao; the first draft of the manuscript was written by Xiayu Tao and guided by Yong Guan, Yangchao Tian, Zhao Wu, and Gang Liu. All authors commented on previous versions of the manuscript. All the authors have read and approved the final version of the manuscript.
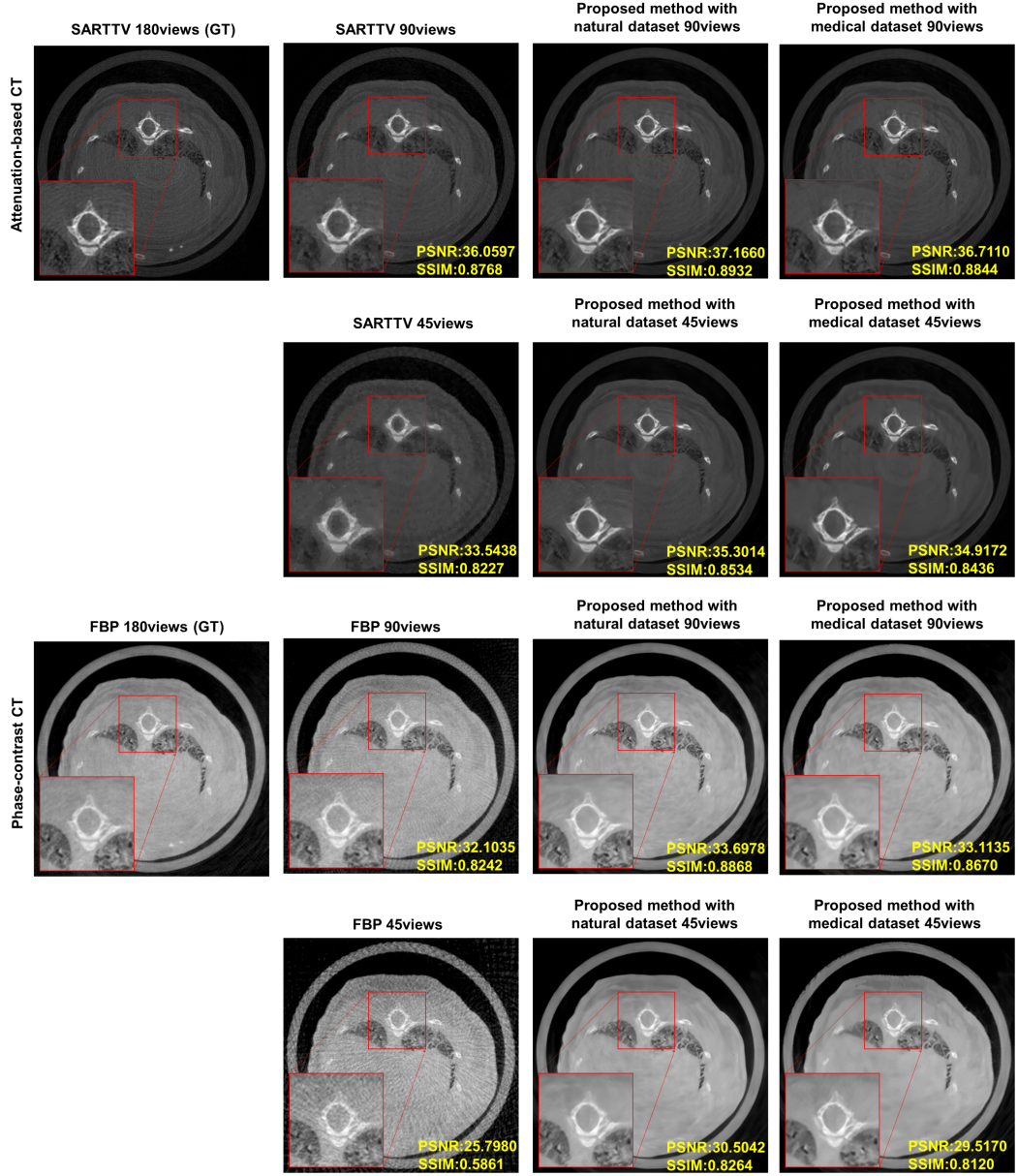
Fig. 4. (Color online) Comparison of reconstructed slices using different methods in mice attenuation-based CT and phase-contrast CT. This figure compares reconstructed slices obtained from various methods in mice attenuation-based CT and phase-contrast CT for phase-contrast imaging. The results shown are derived from the same slice of mice for both attenuation-based CT and phase-contrast CT. In attenuation-based CT, the GT image is reconstructed using the SART-TV algorithm from a set of 180 views. In phase-contrast CT, the GT image is reconstructed using the FBP algorithm from a set of 180 views. We then compared the GT images with the results of the traditional algorithm, the proposed algorithm trained on the natural dataset, and the proposed algorithm trained on the medical data. The comparison was conducted for both 90 and 45 views projection data. The SSIM and PSNR metrics with respect to the GT are displayed in the lower-left corner of each panel. Regions of interest are magnified and indicated by red boxes in the figure. The display range for all results is normalized to [0,255] HU.

SUPPLEMENTARY INFORMATION

A. Performance comparison of the network trained by natural and medical data

To assess the performance of the proposed method, we utilized full-dose human lung slice images for the simulated reconstruction. In this paper, we present a comparison of the proposed algorithm with the original FBP, SART, and SART-TV algorithms in terms of their reconstruction performance under 90 and 45 views of sparse acquisition conditions. The SART and SART-TV algorithms were iterated 150 times to achieve optimal iteration performance. We replaced the sim-
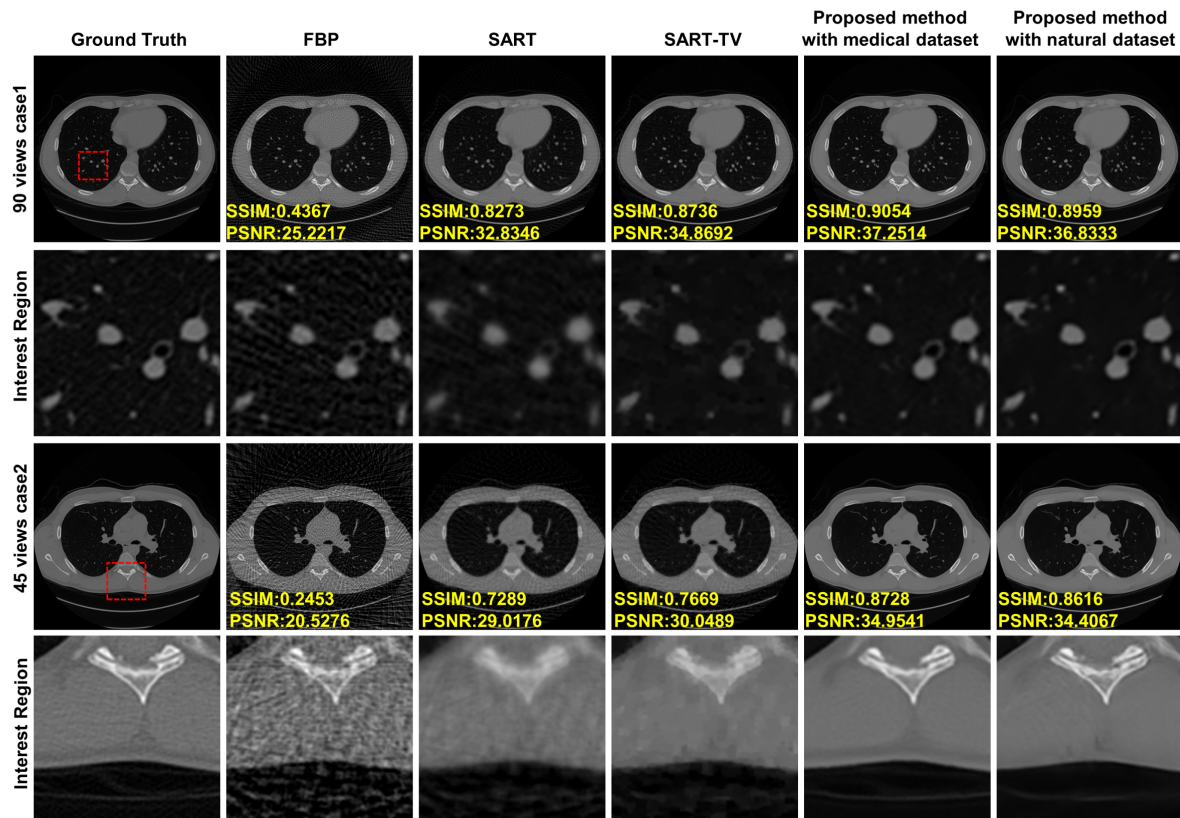
Fig. 5. (Color online) Comparison of reconstructed slices by different methods with medical lung data. The GT images of case 1 and case 2 were obtained from the full-dose slice images of the AAPM dataset, which were reconstructed using 1152 views projected over 180 degrees. The methods used for image reconstruction included Filtered back-projection (FBP), SART, and SART-TV, the proposed method with the medical dataset for training and with the natural dataset for training. These methods reconstructed images from the same 90 views and 45 view projections by forward projecting the GT in parallel geometry. For each reconstruction, SSIM and PSNR metrics with the GT were included in the bottom left corner of the panel. The images of the regions of interest were the zoomed-in areas marked by a red box in the GT. The display range for all results was normalized to [0,255] HU.

ulated dataset with a medical dataset for training and comparison. From the interest-region image of Case 1 in Figure 5, it is evident that the FBP algorithm produced prominent streaking artifacts in the reconstructed image with 90 views. Although the SART effectively suppresses artifacts, the resulting image inevitably suffers from blurring. The SART-TV algorithm preserves the image edges while suppressing artifacts. From the interest region image of Case 2 in Figure 5, we observe that the SART-TV algorithm effectively reduces streaking artifacts caused by sparse-view projection. However, a discernible disparity remained between the SART-TV reconstructed image and the GT image, particularly in the edge regions that exhibited some blurring. Our proposed algorithm removed artifacts better than the SART-TV algorithm and achieved higher scores on image quality assessment metrics. The proposed method using a natural dataset achieved results comparable to those of training on a medical dataset. Our proposed algorithm not only removes artifacts effectively but also preserves the edges of the image consistently with the GT image, without any blurring. This indicates the potential of the proposed algorithm to address the limitations of traditional algorithms and enhance the accuracy and quality of medical image reconstruction. Additionally, our study demonstrates the feasibility of using simulated data from nonmedical datasets to train DL algorithms for medical image analysis, which can facilitate the application of DL techniques in the medical domain with limited availability of medical data.

VI. REFERENCES

[1] T.M. Buzug, Computed Tomography. in the Springer Handbook of Medical Technology. R. Kramme and K.-P. Hoffmann, R. S., Pozos, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), pp. 311–342. doi:10.1007/978-3-540-74658-4_16

[2] H. Kasban, M. El-bendary, and D. Salama, A Comparative Study of Medical Imaging Techniques. International J. Inf. Sci. and Intelligent Systems 4, 37–58 (2015).

[3] P. Rajiah and P. Schoenhagen, The role of computed tomography in pre-procedural planning of Cardiovasc. Surg. and intervention. Insight Imaging **4**, 671–689 (2013). doi:10.1007/s13244-013-0270-8

[4] Z. Wu, W.B. Wei, K. Gao et al., Prototype system of non-interferometric phase-contrast computed tomography utilizing medical imaging components. J. Appl. Phys. **129**, 074901 (2021). doi:10.1063/5.0031392

[5] Z. Wang, N. Hauser, G. Singer et al., Non-invasive classification of microcalcifications with phase-contrast X-ray mammography. Nat. Commun. **5**, 3797 (2014). doi:10.1038/ncomms4797

[6] X. Li, H. Gao, Z. Chen et al., Diagnosis of Breast Cancer-Tokyo based on microcalcifications using grating-based phase contrast CT. Eur. Radiol. **28**, 3742–3750 (2018). doi:10.1007/s00330-017-5158-4

[7] A. Aminzadeh, B.D. Arhatari, A. Maksimenko et al., Imaging Breast Microcalcifications Using Dark-Field Signal in Propagation-Based Phase-Contrast Tomography. IEEE T. Med. Imaging **41**, 2980–2990 (2022). doi:10.1109/TMI.2022.3175924

[8] P. Baran, S. Mayo, M. McCormack et al., High-Resolution X-Ray Phase-Contrast 3-D Imaging of Breast Tissue Specimens as a Possible Adjunct to Histopathology. IEEE T. Med. Imaging **37**, 2642–2650 (2018). doi:10.1109/TMI.2018.2845905

[9] M. Mercuri, T. Sheth, and M.K. Natarajan, Radiation exposure from medical imaging: a silent harm?. Cmaj **183**, 413–414 (2011). doi:10.1503/cmaj.101885

[10] E. Burgio, P. Piscitelli, and L. Migliore, Ionizing Radiation and Human Health: Reviewing Models of Exposure and Mechanisms of Cellular Damage. Epigenetic Perspectives. Int. J. Env. Res. Pub. He. **15**, 1971 (2018). doi:10.3390/ijerph15091971

[11] H.M. Shi, Z.C. Sun, and F.H. Ju, Understanding the harm of low-dose computed tomography radiation to the body (Review). Exp. Ther. Med. **24**, 534 (2022). doi:10.3892/etm.2022.11461

[12] M.D. Cohen, CT radiation dose reduction: can we do harm by doing good?. Pediatr Radiol **42**, 397–398 (2012). doi:10.1007/s00247-011-2315-9

[13] D. Chapman, W. Thomlinson, R.E. Johnston et al., Diffraction enhanced x-ray imaging. Phys. Med. Biol. **42**, 2015–2025 (1997). doi:10.1088/0031-9155/42/11/001

[14] Z. Zhu, K. Wahid, P. Babyn et al., Improved compressed sensing-based algorithm for sparse-view CT image reconstruction. Comput. Math. Methods Med. **2013**, 185750 (2013). doi:10.1155/2013/185750

[15] D. Fleischmann and F.E. Boas, Computed tomography–old ideas and new technology. Eur. Radiol. **21**, 510–517 (2011). doi:10.1007/s00330-011-2056-z

[16] M.J. Willemink and P.B. Noël, The evolution of image reconstruction for CT-from filtered back projection to Artif. Intell.

Eur. Radiol. **29**, 2185–2195 (2019). doi:10.1007/s00330-018-5810-7

[17] A.H. Andersen and A.C. Kak, Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. Ultrasonic Imaging **6**, 81–94 (1984). doi:10.1016/0161-7346(84)90008-7

[18] J. Gregor and T. Benson, Computational analysis and improvement of SIRT. IEEE Trans. Med. Imaging **27**, 918–924 (2008). doi:10.1109/tmi.2008.923696

[19] G. Zang, M. Aly, R. Idoughi et al., Super-Resolution and Sparse View CT Reconstruction. at the European Conference on Computer Vision 2018). doi:10.1007/978-3-030-01270-0_9

[20] D.L. Donoho, Compressed sensing. IEEE T. Inform. Theory **52**, 1289–1306 (2006). doi:10.1109/TIT.2006.871582

[21] Q. Xu, H. Yu, X. Mou et al., Low-dose X-ray CT reconstruction via dictionary learning. IEEE Trans. Med. Imaging **31**, 1682–1697 (2012). doi:10.1109/tmi.2012.2195669

[22] Z. Zhang, X. Liang, X. Dong et al., A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution. IEEE Trans. Med. Imaging **37**, 1407–1417 (2018). doi:10.1109/tmi.2018.2823338

[23] E.Y. Sidky and X. Pan, Accurate image reconstruction in circular cone-beam computed tomography by total variation minimization: a preliminary investigation. in 2006 IEEE Nucl Sci Conf R., 2006), 2904–2907. doi:10.1109/NSSMIC.2006.356484

[24] L.-Z. Deng, P. He, S.-H. Jiang et al.: Hybrid reconstruction algorithm for computed tomography based on diagonal total variation. Nucl. Sci. Tech. **29**, 45 (2018). doi:10.1007/s41365-018-0376-2

[25] S. Niu, Y. Gao, Z. Bian et al., Sparse-view x-ray CT reconstruction via total generalized variation regularization. Phys. Med. Biol. **59**, 2997–3017 (2014). doi:10.1088/0031-9155/59/12/2997

[26] E.Y. Sidky and X. Pan, Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Physics in Medicine Biology **53**, 4777 (2008).

[27] K. Deng, C. Sun, W. Gong et al., A Limited-View CT Reconstruction Framework Based on Hybrid Domains and Spatial Correlation. for sensors (2022). doi:10.3390/s22041446

[28] F. Zhang, M. Zhang, B. Qin et al., REDAEP: Robust and Enhanced Denoising Autoencoding Prior for Sparse-View CT Reconstruction. IEEE Transactions on Radiation and Plasma Medical Sciences **5**, 108–119 (2021). doi:10.1109/TRPMS.2020.2989634

[29] Z. Tian, X. Jia, K. Yuan et al., Low-dose CT reconstruction via edge-preserving total variation regularization. Phys. Med. Biol. **56**, 5949–5967 (2011). doi:10.1088/0031-9155/56/18/011

[30] Y. Liu, J. Ma, Y. Fan et al., Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. Phys. Med. Biol. **57**, 7923–7956 (2012). doi:10.1088/0031-9155/57/23/7923

[31] K. Zhang, W. Zuo, Y. Chen et al., Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Trans. Image Process **26**, 3142–3155 (2017). doi:10.1109/tip.2017.2662206

[32] Y.-J. Ma Y, Ren, P. Feng et al.: Sinogram denoising via attention residual dense convolutional neural network for low-dose computed tomography. Nucl. Sci. Tech. **32**, 41 (2021). doi:10.1007/s41365-021-00874-2

[33] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. (2015). doi:arxiv.org/abs/1409.1556

[34] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. in Medical Image Computing and Computer-Assisted Intervention – MIC-CAI 2015, (Springer International Publishing, 2015), 234–241.

[35] D. Pathak, P. Krähenbühl, J. Donahue et al., and A. A. Efros, Context Encoders: Feature Learning by Inpainting. in 2016 Proc Cvpr IEEE (CVPR), 2016), 2536–2544. doi:10.1109/CVPR.2016.278

[36] C. Ledig, L. Theis, F. Huszár et al., Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. in 2017 Proc Cvpr IEEE (CVPR), 2017), 105–114. doi:10.1109/CVPR.2017.19

[37] H.-K. Yang, K.-C. Liang, K.-J. Kang et al. performed slice-wise reconstruction for low-dose cone-beam CT using a deep residual convolutional neural network. Nucl. Sci. Tech. **30**, 59 (2019). doi:10.1007/s41365-019-0581-7

[38] S.-Y. Zhang, Z.-X. Wang, H.-B. Yang et al.: Hformer: Highly efficient vision transformer for low-dose CT denoising. Nucl. Sci. Tech. **34**, 61 (2023). doi:10.1007/s41365-023-01208-0

[39] J. Kyong Hwan, M.T. McCann, E. Froustey et al., Deep Convolutional Neural Network for Inverse Probl. in Imaging. IEEE Trans Image Process **26**, 4509–4522 (2017). doi:10.1109/tip.2017.2713099

[40] Y. Han, J.J. Yoo, and J.C. Ye, Deep Residual Learning for Compressed Sensing CT Reconstruction via Persistent Homology Analysis. CoRR abs/1611.06391(2016). doi:10.48550/arXiv.1611.06391

[41] Y. Han and J.C. Ye, Framing U-Net via Deep Convolutional Framelets: Application to Sparse-View CT. IEEE Trans. Med. Imaging **37**, 1418–1429 (2018). doi:10.1109/tmi.2018.2823768

[42] S. Guan, A.A. Khan, S. Sikdar et al., Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal. IEEE J. Biomed. Health **24**, 568–576 (2020). doi:10.1109/JBHI.2019.2912935

[43] S. Asif, X. Yongshun, M. Olivia et al., A data generation pipeline for cardiac vessel segmentation and motion artifact grading. in Proc.SPIE, 2022), 122421J. doi:10.1117/12.2642869

[44] Q. Lyu and G. Wang, Conversion Between CT and MRI Images Using Diffusion and Score-Matching Models. ArXiv abs/2209.12104(2022). doi:10.48550/arXiv.2209.12104

[45] S. Misra, C. Yoon, K.-J. Kim et al. developed a deep learning-based multimodal fusion network for the segmentation and classification of breast Cancer-Tokyos using B-mode and elastography ultrasound images. Bioengineering and Translational Medicine **8**, e10480 (2023). doi:10.1002/btm2.10480

[46] A. Zhong, B. Li, N. Luo et al., Image Restoration for Low-Dose CT via Transfer Learning and Residual Network. IEEE Access **8**, 112078–112091 (2020). doi:10.1109/ACCESS.2020.3002534

[47] X. Zhen, J. Chen, Z. Zhong et al., Deep convolutional neural network with transfer learning for rectum toxity prediction in cervical cancer radiotherapy: a feasibility study. Physics in Medicine Biology **62**, 8246 (2017). doi:10.1088/1361-6560/aa8d09

[48] X. Tao, Z. Dang, Y. Zheng et al., Limited-angle artifacts removal and jitter correction in soft x-ray tomography via physical model-driven deep learning. Appl. Phys. Lett. 123, 191101 (2023). doi:10.1063/5.0167956

[49] T.-Y. Lin, M., Maire, S. J., Belongie et al.. Microsoft COCO: Common Objects in Context. at the European Conference on Computer Vision 2014). doi:10.1007/978-3-319-10602-1_48

[50] H.D. Barth, M.E. Launey, A.A. MacDowell et al., On the effect of X-ray irradiation on the deformation and fracture behavior of human cortical bone. Bone **46**, 1475–1485 (2010). doi:10.1016/j.bone.2010.02.025

[51] H. Yu and G. Wang, Finite detector based projection model for high spatial resolution. J Xray Sci. Technol. **20**, 229–238 (2012). doi:10.3233/xst-2012-0331

[52] Z. Wu, K. Gao, Z. Wang et al., Generalized reverse projection method for grating-based phase tomography. J. Synchrotron Radiat. **28**, 854–863 (2021). doi:10.1107/s1600577521001806

[53] Z.-F. Huang, K.-J. Kang, Li Z, et al.: Direct computed tomography reconstruction for directional-derivative projections of computed tomography of diffraction-enhanced imaging. Appl. Phys. Lett. **89**, 041124 (2006). doi:10.1063/1.2219405

[54] K. He, X. Zhang, S. Ren et al., Deep Residual Learning for Image Recognition. in 2016 Proc Cvpr IEEE (CVPR), 2016), 770–778. doi:10.1109/CVPR.2016.90

[55] J. Johnson, A. Alahi, and L. Fei-Fei, Perceptual Losses for Real-Time Style Transfer and Super-Resolution. in Computer Vision (ECCV 2016; Springer International Publishing, 2016), 694–711.

[56] D.P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization. CoRR abs/1412.6980(2014). doi:10.48550/arXiv.1412.6980

[57] A. Paszke, S. Gross, F. Massa et al., PyTorch: An Imperative Style, High-Performance Deep Learning Library. ArXiv abs/1912.01703(2019). doi:10.48550/arXiv.1912.01703

[58] T.R. Moen, B. Chen, D.R. Holmes et al., Low-dose CT image and projection dataset. Med. Phys. **48**, 902–911 (2021). doi:10.1002/mp.14594

[59] X. Ge, P. Yang, Z. Wu et al., Virtual differential phase-contrast and dark-field imaging of x-ray absorption images via deep learning. Bioengineering Translational Medicine **8**, e10494 (2023). doi:10.1002/btm2.10494